Volume 15 Number 6 1987 Nucleic Acids Research

The length but not the sequence of the polyoma virus late leader exon is important for both late RNA splicing and stability

Guy R.Adami§ and Gordon G.Carmichael

Department of Microbiology, University of Connecticut Health Center, Farmington, CT 06032, USA

Received July 8, 1986; Revised and Accepted January 27, 1987

ABSTRACT

Polyoma virus late RNA processing provides a convenient model system in which to study the mechanics of splicing *in vivo*. In order to understand further the role of the untranslated "late leader" unit in late RNA processing we have constructed a group of polyoma viruses with deletions and substitutions in the leader exon. This has allowed us to determine that there is a minimum exon size required for both pre-mRNA splicing and stability in this system. We show here that the non-viability of a mutant (ALM) with a 9 base late leader unit is due to a general defect in late RNA splicing. In addition, ALM-infected cells show at least 40-fold depression in the accumulation of late nuclear RNA (spliced or unspliced). The ALM late promoter, however, functions nearly normally. Substituted leader variants with 51- to 96-base long exons of unrelated sequence are viable (G. Adami and G. Carmichael, J. Virol. 58, 417-425, 1986). We show here that late RNA from one of these substituted leader mutants (containing a 51-base leader exon) is spliced at wild type levels, with virtually no defect in accumulation. Thus, in the polyoma system, splice sites separated by only 9 bases can inhibit each others usage, presumably by steric interference. We suggest that this type of inhibition leads to extreme RNA instability.

INTRODUCTION

Much work in recent years has been directed towards further understanding the mechanism of splicing of mRNA precursors. Analyses of intron sequence requirements for *in vivo* mRNA splicing in higher eukaryotes originally indicated the importance of the 5' splice site, including the highly conserved GU dinucleotide at the 5' exon-intron boundary, and the 3' splice site, which consists of a similarly highly conserved dinucleotide, AG, at the 3' boundary, but which additionally includes an upstream stretch of pyrimidines (1-7). More recently, a third site which shows some sequence requirements for splicing *in vivo* has been identified - the branch point, which lies about 30 nucleotides upstream of the 3' splice site (8). Results with deletion mutants and chimeric RNA's have indicated little or no sequence requirements for exon sequences and intron regions not already mentioned (7,9,10) - these other regions may only have a role in determining three-dimensional structures which facilitate the bringing together of companion splice sites as originally proposed (11-13).

Work with *in vitro* splicing systems has also made possible the identification of pre-mRNA sites that interact with the splicing machinery. It has been known for some time that the same three sites which show sequence requirements for efficient splicing are also the sites on the precursor RNA that are active in splicing - the sites of 5' intron cleavage, lariat formation and 3'

intron cleavage. All three of these regions (and these alone) have been characterized as binding sites for components of the splicing machinery during *in vitro* splicing. This has been done using several different approaches. Steitz's laboratory (14-16) has used immunoprecipitation methods to demonstrate the association of the 5' splice site with U1 snRNP particles, the branch point with U2 snRNPs, and the 3' splice site with U5 snRNPs. Interestingly, binding of factor(s) to the 3' splice site (the 19 bases in the intron preceding the 3' splice site) requires no other sequences. Ruskin and Green (17) also investigated binding sites in an *in vitro* splicing reaction. They determined which pre-mRNA sites are bound by factors and thereby lose their ability to hybridize to short complementary oligonucleotides as measured by resistance to RNAse H. These experiments revealed binding sites similar to those reported by Steitz's group, but with a distinct hierarchy of binding. Protection at the branch point (dependent on an intact 3' splice site) was observed first. Later in the lag period during *in vitro* splicing, but prior to splicing, protection of the 5' splice site was seen.

These data are brought into perspective by the identification of a multicomponent splicing complex. The splicing complex shows the same kinetics of formation as do spliced products after mixing of mRNA precursor and splicing extract. It is likely that the binding sites characterized by the laboratories of Green and Steitz are brought together under the direction of a single splicing complex. Rigorous analysis of the sequences required for "spliceosome" formation has indicated that the 5' and 3' conserved sequences are both necessary for complete 50S splicing complex formation (18). An intermediate splicing complex forms with the RNA substrate when the 3' splice site is present but the 5' splice site is deleted (18). These and other results support a model where splicing occurs in a multifactor complex with a hierarchy of binding to the three conserved sites on the precursor mRNA.

We have studied splicing of polyoma virus late pre-mRNA molecules. Late transcription termination in polyoma is inefficient (19). As shown in Figure 1A, many polyoma late primary transcripts contain tandem repeats of the entire genome, caused by multiple circuits of the viral chromosome by RNA polymerase II. Thus, the 57-base late leader exon, which is encoded only once in the viral genome, can be present multiple times in late pre-mRNA molecules. Late cytoplasmic polyoma mRNA contains direct repeats of these leader exons presumably generated by efficient leader-to-leader splicing (20). We have previously described a polyoma mutant, ALM with an abbreviated leader exon of 9 nucleotides. ALM is nonviable and is at least 200-fold deficient in late cytoplasmic mRNA accumulation but is unaffected in early gene expression (21). Restoration of leader length with unrelated sequences can restore viability (21). The observed deficiency in ALM cytoplasmic late viral mRNA levels made it imperative to determine whether the defect lay in the late viral promoter, or in late RNA splicing, transport or stability. We establish here that ALM has a general defect in late RNA splicing, although all introns and splice sites are intact. In addition, the late RNA that is made in ALM is highly unstable. Restoration of leader length with an unrelated sequence relieves both defects. These results provide evidence that

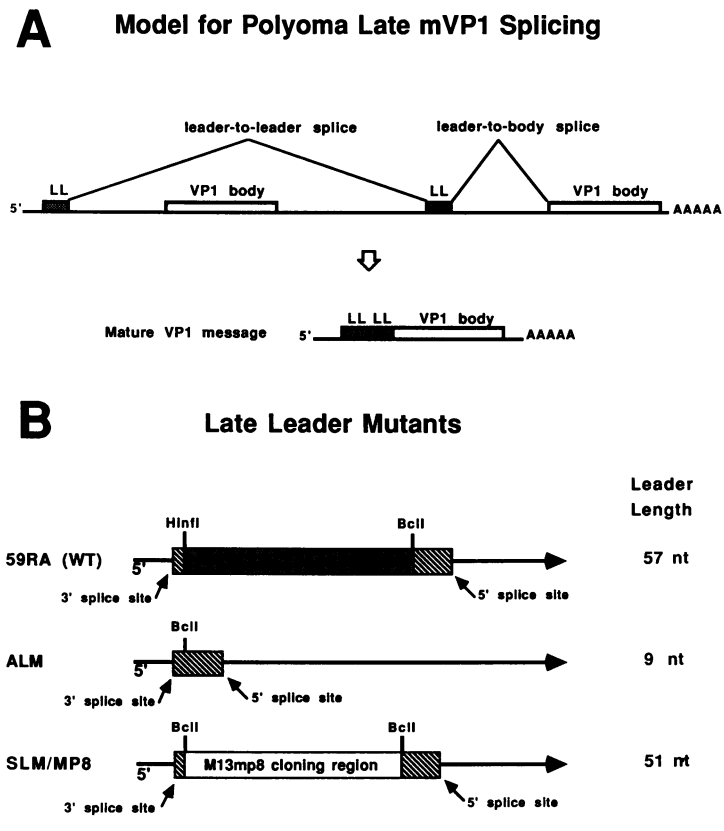

Figure 1. A. A model for the processing of polyoma giant late primary transcripts and the production of multiple, tandem non-translated leader units on mature mRNA molecules. Due to inefficient termination by RNA polymerase II late viral transcripts are heterogeneous in length, with many representing multiple circuits of the genome by the polymerase. The example shown here is for a transcript created by two circuits by RNA polymerase II. In the first step of processing, a message body (here, mVP1) is spliced to the 3'-proximal late leader unit (LL). Then, leader-to-leader splicing removes genome length introns and yields the final mVP1 message which contains 2 tandem leader units at its 5'-end. B. Schematic diagram of the late leader regions of the viruses used in this study. The heavy lines represent late RNA; the boxed areas are the leader units. Stippled areas represent leader sequences retained by all three viruses. Our wild type, 59RA, has a 57-base late leader unit about 50-80 nucleotides downstream of the major late startsites. There are no AUG codons in this region. The nonviable mutant ALM (21) was created by removing 48 nucleotides from 59RA from a HinfI site to a BclI site. SLM/MP8 (21) was made by inserting the multi-site cloning region of M13mp8 into the unique BclI site retained by ALM.

binding sites characterized as necessary for splicing can inhibit each others' usage, and support earlier speculation that juxtaposition of splice sites leads to RNA instability (7). Together these results support the notion that splicing complex formation is involved in nuclear pre-mRNA preservation, at least in the polyoma system.

MATERIALS AND METHODS

Materials

The plasmids used here, pPy (containing the entire wild type polyoma genome), pALM (containing the genome of the abbreviated leader mutant) and pSLM/MP8R (containing the genome of a substituted leader mutant), were constructed and propagated as described previously (21), and are diagrammed in Figure 1B. For simplicity, the virus referred to in this paper as "SLM/MP8" is the same as the virus "SLM/MP8R" in our previous manuscript (21). Random hexanucleotides for oligonucleotide labeling of DNA probes were made using a Systec Model 1450 DNA Synthesizer. Restriction enzymes were from New England Biolabs and were used as suggested by the supplier. S1 nuclease was from Sigma. T4 polynucleotide kinase, DNA ligase and *E. coli* DNA polymerase I, large fragment were from New England Biolabs. [α-^{32}P]-deoxynucleoside triphosphates and [γ-^{32}P]-ATP were from New England Nuclear or Amersham. All other reagents and enzymes were from commercial suppliers.

Cell culture techniques

Mouse NIH3T3 cells were grown and transfected with viral DNAs or infected with virus as described previously (21). At 40 or 42 hours after transfection nuclear RNA was isolated (22) and stored as an ethanol precipitate at -70° until used for S1 analysis.

Transcription rate measurement in isolated nuclei

Thirty hours post transfection nuclei from mock-transfected, ALM-transfected or wild type (pPy)-transfected cells were prepared, incubated with [α-^{32}P]-UTP, and the labeled RNA isolated, and hybridized to single stranded probes bound to nitrocellulose filters. Nuclei were prepared and transcription reactions were done as described (22). For each experiment, nuclei were used from two 90 mm plates of transfected NIH3T3 cells. Following incubation, the reactions were terminated with DNAse I, and RNA was isolated essentially as described (23), except that instead of trichloroacetic acid precipitation of RNA and collection onto filters, LiCl precipitation followed by centrifugation (24) was used. RNA fragmentation (22) was followed by addition of carrier yeast RNA. Hybridization to single stranded (polyoma-M13 recombinant phage) DNA probes specific for early or late transcripts bound to nitrocellulose was as described (22) except a 2 hour preincubation was used. Single strand DNA was spotted using a BRL slot-blot manifold apparatus to allow quantitation of the autoradiograms by densitometry. Five micrograms of DNA were bound per slot.

Preparation of probes and S1 analysis

Most S1 experiments used short, end-labeled single-stranded probes (25). Plasmid pPy DNA was digested with appropriate restriction enzymes and when necessary the desired fragments were isolated by polyacrylamide gel electrophoresis, electroelution (using an International Biotechnologies apparatus), and ethanol precipitation. The 5'-ends were dephosphorylated using calf intestine alkaline phosphatase, purified further using an elutip-d column (Schleicher and Schuell) and then labeled with polynucleotide kinase and [γ-^{32}P]-ATP

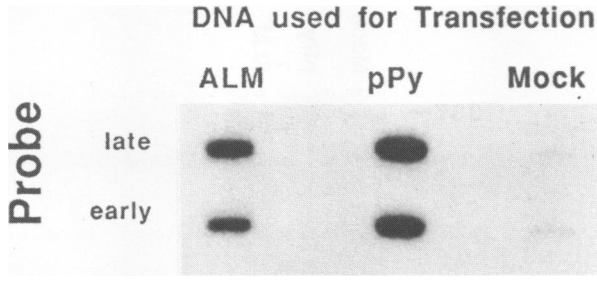

Figure 2. Nuclear runoff assay to measure late promoter strength in ALM. Viral genomes propagated as BamH1 inserts into plasmids were excised and recircularized by ligation under dilute conditions and used to transfect mouse NIH3T3 cells. Transcription rate measurements were as described in Materials and Methods. Early and late specific probes were viral BamH1-EcoR1 fragments cloned into phage M13mp8 (binds late RNAs) or M13mp9 (binds early RNAs). Since ALM early gene expression is indistinguishable from wild type (21), the early probes here were used as an internal control. The efficiency of late promoter function was determined by densitometric scanning of the autoradiograms and comparison of the relative levels of signal with the early and late probes.

(26). The strands were separated using polyacrylamide-urea gel electrophoresis (27), and electroeluted from the gels.

RNA for analysis (generally 1/2 to 1/50 of the RNA isolated from 1 plate) was mixed with labeled probe (0.1-1μg) and hybridization and S1 digestion were as described (26), except 70% formamide was used, and annealing was at 36°C. Optimal temperatures for the 1 hour incubation with nuclease varied depending on the probe used, and are mentioned in specific figure legends. The products were resolved on 6% polyacrylamide-urea sequencing gels (27). Quantitation was by laser densitometry of the autoradiograms. To ensure that S1 analysis was always done in probe excess, varying amounts of probe were used in other experiments.

Long single-stranded DNA probes were made by cloning a 3072 base pair polyoma EcoR1-BamH1 fragment into M13mp8 or M13mp9 (28). Single strand DNA was isolated from purified recombinant phages (28) and used for S1 analysis just as described above for short, end-labeled probes. Following digestion of hybrids, the products were fractionated on 2% agarose gels in 40 mM Tris-acetate, pH 7.8, 1 mM EDTA at 5-20 V/cm until the bromophenol dye migrated 8 cm. Bands were transferred to GeneScreen Plus (New England Nuclear) and hybridized using the manufacturer's instructions to a ^{32}P-labeled polyoma-specific probe prepared by the oligonucleotide labeling technique (29).

RESULTS

Efficient late transcription does not require the late leader sequence

The late promoter of polyoma is not well characterized. We considered it possible that the defect in ALM late RNA accumulation is due to a promoter defect, although the leader lies

Figure 3. ALM and wild type viruses show the same pattern of late RNA 5'-ends. Total nuclear RNA was isolated from mock-, pPy (wild type)-, or ALM-transfected cells and analyzed using S1 and a single-stranded 337 nucleotide probe (HinfI - BglI) which spans the late 5'-ends and the replication origin. S1 digestion was carried out at 15°C. Since ALM accumulates only low levels of late RNA, much more ALM RNA than pPy RNA as isolated had to be used for this experiment. Quantitation of these data required a transfection efficiency control, which was an analysis of early RNA levels (Figure 4). The prominent band at about 85 nucleotides represents a major late startsite, and the larger bands represent minor startsites. The 337 nucleotide band reflects precursor molecules. Band sizes were determined using end-labeled pBR322 HinfI markers (not shown).

downstream of most or all late RNA start sites used *in vivo* (20,30,31). The late leader might serve to separate as yet unidentified upstream and downstream promoter elements so they can function efficiently. Therefore, a nuclear run-off transcription assay was done. Nuclei were isolated from transfected cells and incubated with [α-^{32}P]-UTP. This allows labeling of preinitiated RNA transcription complexes and provides a kinetic analysis of RNA transcription levels. Using two single-stranded DNA probes, one that specifically binds to late RNA and one that binds to early RNA, levels of transcription were examined in ALM and pPy (Figure 2). Using levels of early transcription as an internal reference the ratios of band intensities indicate that ALM late transcription is indistinguishable from that of pPy. Unexpectedly, for both ALM and pPy late transcription as measured by the probes used is about twice as efficient as early transcription. When kinetic analysis of transcription levels is done post virus infection late transcription dominates by a factor of 5-10 (32). This has been verified in our laboratory using the nuclear transcription assay and the same DNA probes as above bound to filters (data not shown). This observation suggests that the temporal control of transcription of polyoma RNA may not be identical post transfection and post infection. Nevertheless, we can conclude that the accumulation defect in ALM is not due to a crippled late viral promoter.

ALM late transcripts show the same pattern of 5'-startsites as wild type virus

To determine whether major new late RNA startsites are being used by the mutant, S1 mapping was done of post-transfection late nuclear RNA using a homologous single strand antisense 5'-^{32}P-HinfI - BglI probe (Figure 3). The 5'-end of this probe lies just inside the leader exons of both ALM and pPy late RNAs and extends upstream, across the viral DNA replication origin. Results in Figure 3 indicate similar patterns of start sites for ALM and pPy late nuclear RNA, although much more total RNA as isolated from ALM-transfected cells had to be used for the analysis. The most prominent 5'-end maps approximately to viral nucleotide 5127 (numbering system as in 33), the major start site detected by other groups (31,34). We should note that our wild type laboratory strain, 59RA, contains a 30 base pair duplication which includes this site (35). We have noticed using longer probes (data not shown) that in this strain each "5127" appears to be used at the same efficiency, resulting in two prominent 5'-ends. The data in Figure 3 show the upstream "5127" startsites. Bands corresponding to mRNA's with start sites at the downstream homologous "5127" site are visible as bands only on longer exposure, since they protect probe fragments of about 50 bases in length which are relatively unstable under the conditions used here for S1 digestion. In addition, longer exposure reveals that the minor late startsites noted by others are also used by both 59RA and ALM. Protection of full length probe in Figure 3 is indicative of unspliced or partially spliced giant pre-mRNA's which contain tandem repeats of the polyoma genome.

Figure 4 shows the results of S1 mapping of early RNA from the same transfections done with a single stranded 5'-^{32}P-AvaI - HgiAI probe which binds across early mRNA splice sites. As ALM has been shown to be wild type in early functions (21) this was used in all experiments

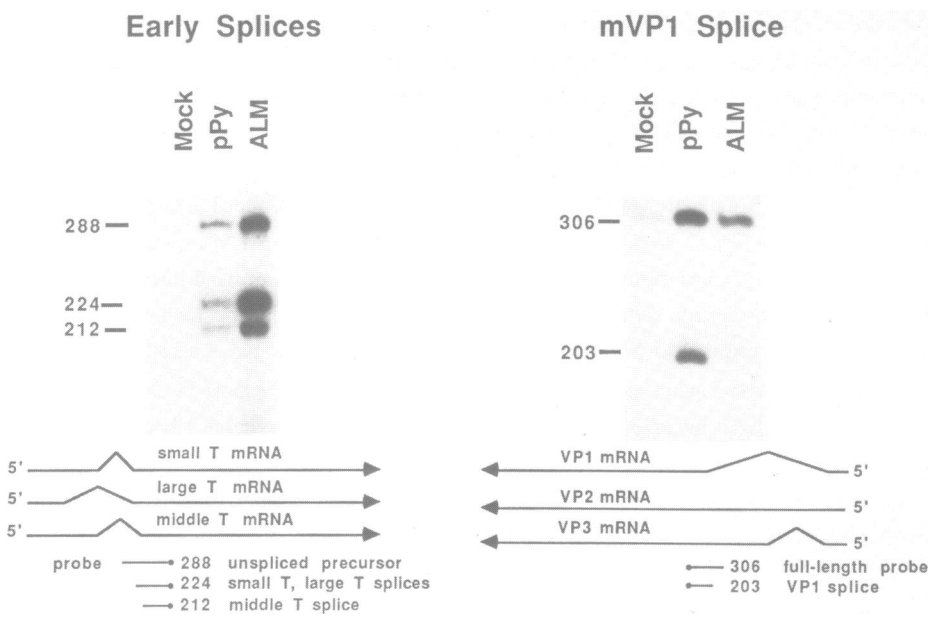

Figure 4. ALM is defective in late nuclear RNA accumulation and mVP1 splicing. Nuclear RNA was isolated from cells 40 hours after transfection and subjected to S1 analysis using a 288 nucleotide early probe (AvaI - HgiA1), which reveals splices for the messages for small, middle and large T antigens, or a 306 nucleotide (HindIII - PstI) late probe, which reveals the mVP1 splice. S1 digestion was at 22°C. The same relative amounts of pPy and ALM RNAs were used for the two experiments shown here to allow quantitation. Densitometric scanning of various exposures of this autoradiogram as well as films from independently repeated experiments consistently suggested a 40-fold defect in ALM late nuclear RNA accumulation using the mVP1 probe and a further at least 10 fold reduction in mVP1 splicing.

as an internal control for transfection efficiencies and RNA recoveries. Densitometry of the bands from both of these experiments revealed that, although the pattern of late start sites detected by the HinfI - BglI probe is as wild type, there is about a 40-fold defect in ALM in the accumulation of late 5'-ends in nuclear RNA. Other experiments using a 1200 nucleotide BamH1-PstI probe (not shown) confirmed that no 5'-ends mapping downstream of the leader unit could be detected for either pPy or ALM. Thus, it appears that transcriptional initiation is at the same sites in wild type and ALM.

ALM has a general defect in late RNA splicing

Due to the small size of the leader exon in ALM we considered it possible that late RNA in this mutant is spliced inefficiently. Analysis of leader-to-leader splicing with only a 9 base leader repeat, however, would not be feasible using S1 mapping. Instead, a single strand 5'-^{32}P-HindIII - PstI probe was used for S1 analysis of leader-to-mVP1 body splicing (Figure 4). A signal apparent at 306 nucleotides would indicate full length protected probe, and a signal at 203

nucleotides would indicate leader-to-body mVP1 splicing. Full length protected probe is indicative of unspliced RNA, partially spliced RNA, or RNA not spliced in the region of the mVP1 3' splice site, and shows at least a 40-fold deficit in accumulation in ALM vs pPy, as determined by densitometry of these autoradiograms and correction for the amounts of early RNA present. More interestingly, there is a marked deficit in the 203 base signal in ALM vs full length protected probe, indicating a further 20-fold deficit in accumulation of RNA with the mVP1 splice.

A different leader-to-body splice is required for mVP3 synthesis. In other S1 experiments we verified that this splice, too, is at least 10-fold depressed in ALM-infected cells (data not shown). Two out of three splices made in the biogenesis of polyoma late mRNAs have thus been tested and shown to be defective. The only late splice not tested was the leader-to-leader splice.

ALM late RNA is correctly processed at the 3'-end, and no major cryptic splice sites are used

Villareal et al. (36) have reported the existence of an SV40 mutant that contains a deletion of a late mRNA intron and splice junctions. Late RNA from this mutant is apparently not efficiently polyadenlyated, nor is it transported to the cytoplasm. It was unlikely that this SV40 mutant is similar to ALM, as it exhibited the accumulation of non-polyadenylated RNA and also showed an altered pattern of late 5'-ends. Nevertheless, to firmly establish that the splicing defect in ALM was the primary lesion, a 3'-end analysis of ALM late nuclear RNA was undertaken.

Quantitative S1 analysis was done with large single stranded DNA probes which would allow detection of the 3'-ends of processed transcripts. Use of large probes would not only rule out the possibility of aberrant 3'-end formation in ALM, but would also allow the detection of cryptic splice sites. Measurement of early RNA levels was used as an internal standard, and much more ALM RNA than pPy RNA was used. Results of this S1 analysis of nuclear RNA are shown in Figure 5. In the short exposure of the autoradiogram the early ALM signal is readily observed, as are the bands representing pPy late processed RNAs. Whereas the wild type late signal is greater than the early one, the opposite holds for ALM. Even in the overexposure, ALM late RNA, spliced or unspliced, is barely detectable, while ALM early RNAs are expressed at high levels. The two ALM late signals indicative of unspliced RNA (full-length probe corresponding to giant precursor RNA, and 1860 base pairs representing unspliced but 3'-end processed RNA) are present in about equal amounts, as is the case for wild type. No band corresponding to the mVP1 splice can be observed for ALM, although it is the dominant band with wild type late nuclear RNA. These results allow us to conclude that 80% of stable nuclear wild type RNA contains the proper processed 3'-ends. In ALM RNA we see qualitatively that properly 3'-end-processed RNA represents about 50% of the total. The results suggest that if there is a defect in cleavage/polyadenylation of ALM late RNA, it is a minor one. Also, they serve to illustrate the severity of the ALM defect in late RNA accumulation. Additionally, this and repeated experiments reveal no prominent usage of cryptic splice sites within 400 nucleotides of the mVP1 splice sites. The possible band at about 1600 nucleotides in the ALM late lane has not been seen

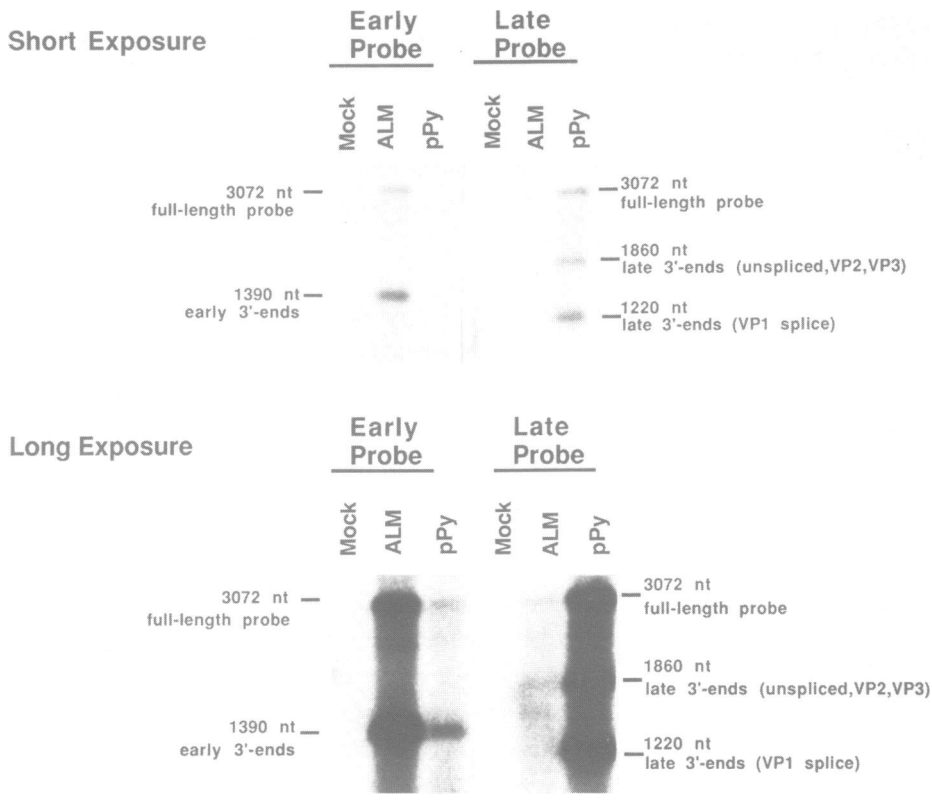

Figure 5. A large S1 probe reveals apparently normal 3'-end processing and a dramatic defect in ALM late nuclear RNA accumulation. A 3072 base pair EcoR1-BamH1 polyoma fragment was cloned into M13mp8 and into M13mp9 RF DNA. These recombinant phages provided single-stranded DNA complementary to early (mp9) or late (mp8) viral RNAs. The single-stranded probes were annealed in excess to nuclear RNA isolated 40 hours after transfection, and the complexes digested with S1 nuclease at 36°C. The same amounts of pPy and ALM RNAs were used for both early and late analyses. Fully protected probe represents precursor RNA molecules. For early RNAs a 1390 nucleotide band represents normal 3'-end processing. For late RNAs, an 1860 nucleotide band represents molecules with normal 3'-end processing but no mVP1 splice. These would include mVP2 and mVP3 molecules, as well as unspliced transcripts. A 1220 nucleotide band represents molecules with correct 3' termini and the mVP1 splice.

in other experiments. Finally, when compared to the levels of early transcripts, these results reveal at least a 400-fold deficit in the accumulation of intact late RNA in ALM-transfected cells.

In an effort to increase sensitivity, an S1 analysis of 3'-ends of late nuclear RNA was done using a short, end-labeled, single strand DNA probe (Figure 6). The probe used was a 236 nucleotide NcoI-HgiA1 fragment. S1 digestion following hybridization to nuclear RNA revealed a protected band of about 165 nucleotides, corresponding to the expected 3'-end (30 nt

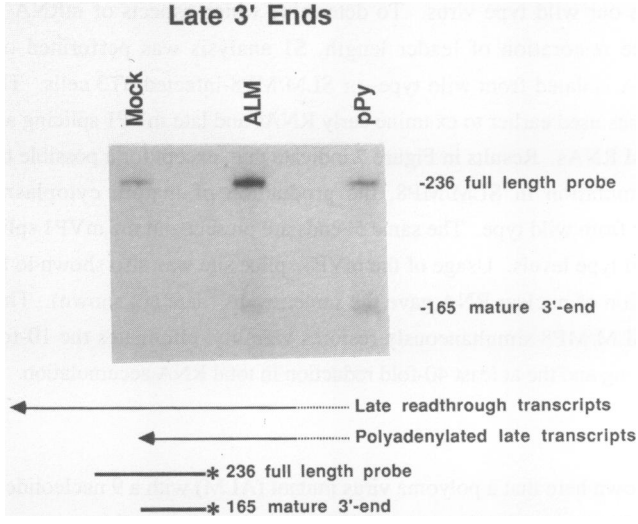

Figure 6. A short S1 probe reveals efficient 3'-end processing by ALM. Plasmid pPy (5 μg) was cut with NcoI and the ends labeled with α-^{32}P-dATP and the Klenow enzyme. Then the DNA was cut with HgiA1 and labeled fragments separated on a 4% acrylamide-7M urea strand-separating gel. The 236 nucleotide band which spans the late polyadenylation site was electroeluted, annealed in excess to nuclear RNA at 37°C, and complexes digested with S1 nuclease at 22°C. For this experiment, comparative quantitation using the mVP1 probe showed that the late ALM processed end (protected band of about 165 nt) is present at about 1/100 of the level of the pPy processed end.

downstream of the AAUAAA sequence). Cleavage at this site in polyoma has been shown to be tightly coupled to polyadenylation (37). Full-length probe is indicative of readthrough transcripts not polyadenylated at the expected site and perhaps due to multi-genomic transcripts. The background band in the mock lane of Figure 6 suggests that our probe is not completely single stranded. When corrected for the background, what is revealed here is that wild type shows a ratio of about 2:1 for mature 3'-ends to readthrough. For ALM this ratio becomes about 1:2. This would indicate that for stable late nuclear RNA that includes the polyA site about a third of ALM late nuclear RNA is properly cleaved and polyadenylated, while in wild type this figure approaches 80%. However, less than 5% of stable RNA in ALM contains the mVP1 splice (Figure 4), while 50% in pPy does. As the splicing defect is the graver of the two, we suggest it is the primary defect. Quantitation reveals that ALM polyadenylated 3'-ends are present at about 1/100 the levels of wild type.

Splicing and stability defects of ALM RNA are linked

A number of substituted leader mutants, with leader exons of 51 to 96 nucleotides in length but unrelated sequence, are viable (21). An examination was done of late mRNA production in cells infected by one of these mutants, SLM/MP8, which contains a 51-base leader unit and

grows as well as our wild type virus. To determine which aspects of mRNA production are alleviated by the restoration of leader length, S1 analysis was performed on nuclear and cytoplasmic RNA isolated from wild type- or SLM/MP8-infected 3T3 cells. The probes used were the same ones used earlier to examine early RNAs and late mVP1 splicing and 5' start sites for pPy and ALM RNAs. Results in Figure 7 indicate that, except for a possible two-fold deficit in mRNA accumulation in SLM/MP8, the production of mature cytoplasmic mRNA is indistinguishable from wild type. The same 5'-ends are present and the mVP1 splice site appears to be used at wild type levels. Usage of the mVP3 splice site was also shown to be as wild type and an examination of nuclear RNA gave the same results (data not shown). Thus the restored leader exon of SLM/MP8 simultaneously restores viability, eliminates the 10-fold (or greater) reduction in splicing and the at least 40-fold reduction in total RNA accumulation.

DISCUSSION

We have shown here that a polyoma virus mutant (ALM) with a 9 nucleotide exon in all late RNA transcripts has a general defect in late RNA splicing *in vivo*. All known coding sequences, splice sites and intron sequences, however, are intact. In addition to the splicing defect, ALM late nuclear RNA is unstable and accumulates to 1/40 or to 1/400 the levels of wild type nuclear RNA's, depending on the method of measurement (see below). The late promoter of ALM functions normally - transcripts initiate at the same sites and levels as in wild type. No new start sites can be detected between 300 bases upstream and 200 bases downstream of the major wild type startsite. Both the splicing defect and the stability defect are relieved by restoring the length of the late leader exon with an unrelated sequence (SLM/MP8, see Figure 7). The defect is unlikely to be due to a shortened distance between 5'-ends and splice sites, since our strain of polyoma contains a 30 bp duplication. Thus in ALM, late startsites are, in fact, positioned almost the same distance from splice sites as for the A2 strain of polyoma (31,34). What the data suggest is that the two defects are related and this is likely the first clear demonstration that juxtaposed 5' and 3' splice sites can interfere with each other's usage resulting both in inhibition of splicing and in RNA instability.

There have been a number of reports which link splicing defects to RNA stability. Large deletions that interfere with correct splicing have been shown to have differing effects on RNA stability (38-40). The magnitude of the deletions compared to the wild type controls used in these studies, coupled with the variable results, makes interpretation of these data difficult. Mutants with smaller deletions that remove splice site consensus sequences have also been studied (41,42). Some of these mutant transcripts do not yield any detectable spliced products. Instead, unspliced precursors accumulate or are degraded. These systems have not been studied rigorously enough to make obvious to us a unifying pattern in the results. An examination of results of analyses with the large group of splicing mutants that contain splice sites inactivated by small deletion and point mutations is more fruitful. In such studies splice site inactivation almost

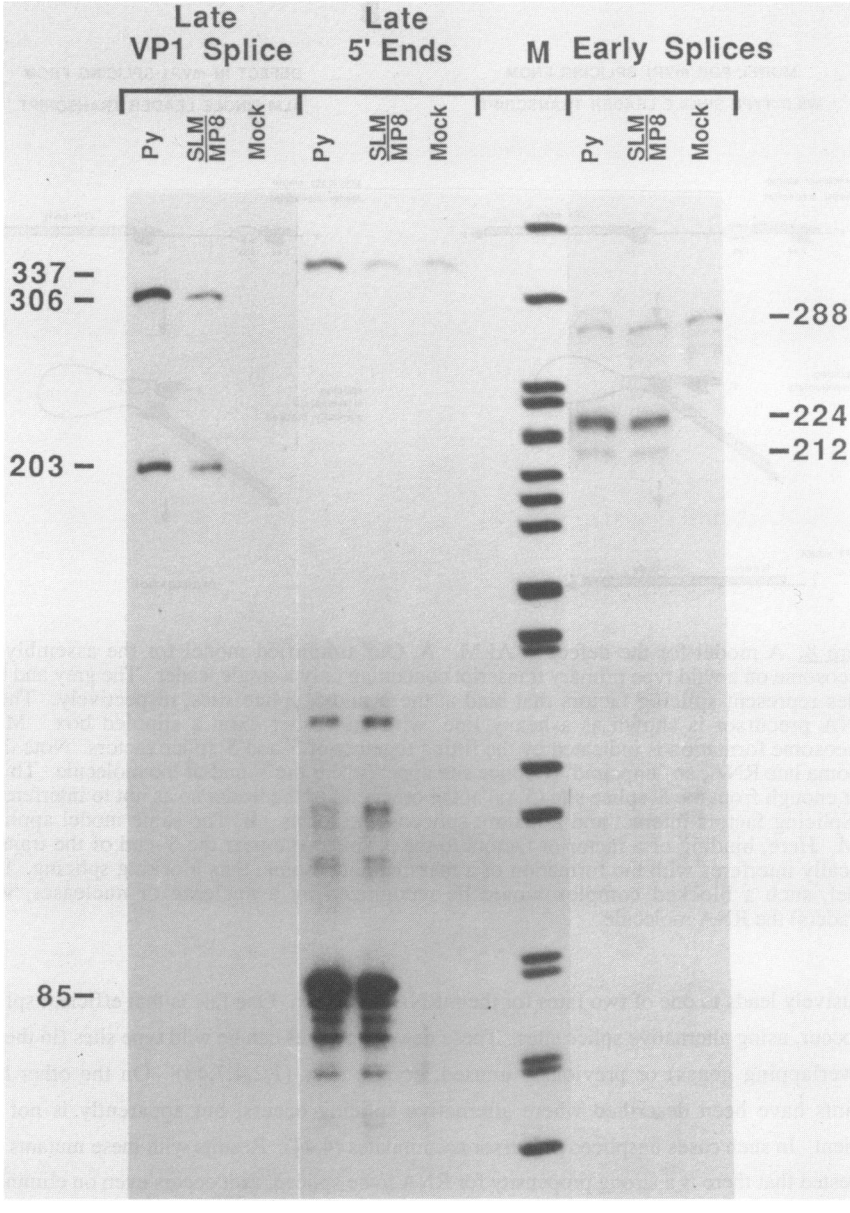

Figure 7. S1 analysis of cytoplasmic SLM/MP8 RNA. Since SLM/MP8 is viable and grows as well as our wild type 59RA (Py), mouse NIH3T3 cells were infected with each virus at equivalent multiplicity of infection. At 40 hrs post infection cytoplasmic RNA was isolated and subjected to S1 analysis using the same probes used earlier for ALM (Figs. 3,4) in order to demonstrate early and late splicing and the pattern of usage of late 5'-startsites. M, pBR322 HinfI size markers.

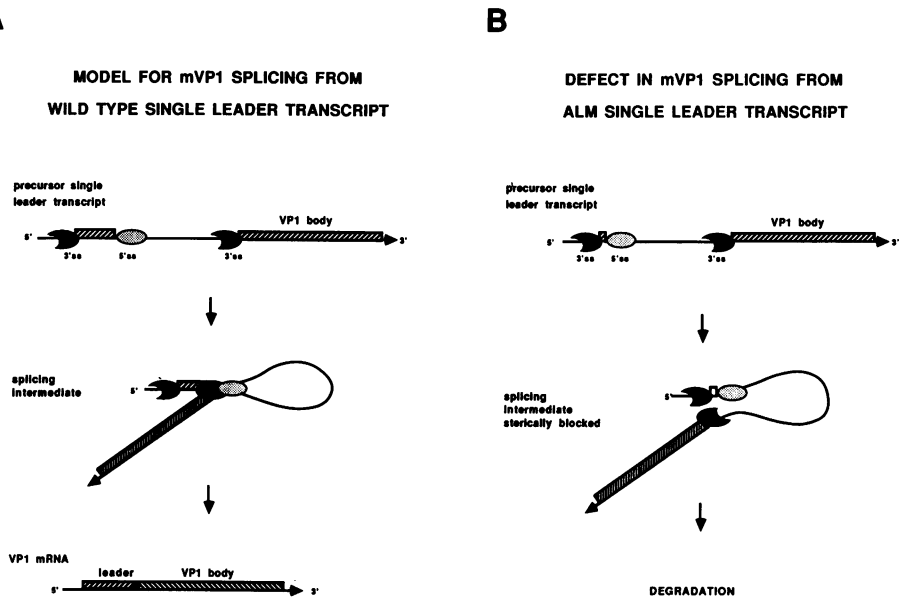

Figure 8. A model for the defect in ALM. **A.** Our simplified model for the assembly of a spliceosome on a wild type primary transcript containing only a single leader. The grey and black shapes represent splicing factors that bind at the 3' and 5' splice sites, respectively. The late mRNA precursor is shown as a heavy line, with the leader exon a stippled box. Mature spliceosome formation is indicated by the fitting together of 5' and 3' splice factors. Note that in polyoma late RNA, an "unpaired" 3' splice site appears near the 5'-end of the molecule. This site is far enough from the 5' splice site (5' ss) at the other end of the leader so as not to interfere with it. Splicing factors interact and a mature spliceosome forms. **B.** The same model applied to ALM. Here, binding of a factor or factors to the 3' splice site near the 5'-end of the transcript sterically interferes with the formation of a mature spliceosome, thus blocking splicing. In our model, such a blocked complex would be recognized by a nuclease or nucleases, which degrade(s) the RNA molecule.

exclusively leads to one of two fates for the mRNA precursor. One fate is that efficient splicing can occur, using alternative splice sites. These new splice sites can be wild type sites (in the case of overlapping genes) or previously unused, cryptic sites (1,2,4,7,43). On the other hand, mutants have been described where alternative splicing occurs, but apparently is not very efficient. In such cases unspliced precursor accumulates (4,44). Results with these mutants have suggested that there is a strong propensity for RNA to be spliced, as it occurs even on elimination of a wild type splice site. They show no ostensible connection between RNA splicing and RNA stability.

However, results with some ß-globin splice site mutants do suggest that our findings with ALM, linking an RNA splicing and stability defect, are not anomalous. When splice site inactivation happens to lead to equivalent usage of closely situated cryptic splice sites splicing is

inefficient and total levels of RNA (spliced and unspliced) are depressed. Treisman et al. (5) examined a 5'-splice site point mutant in human ß-globin IVS1 where splicing occurs at three cryptic sites. All of these new sites lie within a stretch of 50 nucleotides and appear to be used equally. In this mutant, RNA accumulation is depressed 10-20 fold, with no unspliced precursor seen. Another point mutation which results in only partial deactivation of the same 5'-splice site still allows this site to dominate over the same three cryptic splice sites, which are used at very low levels. Here no deficit in RNA is seen. Wieringa et al. (7) have reported results where crowded splice sites of equal activity (or efficiency) seem to inhibit each others' usage resulting in low levels of accumulation of the particular RNA. This group examined a rabbit ß-globin IVS2 derivative with an altered 3'-splice site. The new splice sites used in this mutant RNA are of similar efficiency (they presumably interact equally well with the splicing machinery) and are separated by only 20 bases. These authors further speculate that when splice sites of similar usage levels are closely situated they can inhibit each others' function and thus levels of spliced product are decreased. Our results with ALM show that a precursor RNA with 9 bases between two efficient splice sites is not spliced well, while one with a 51 base separation is. These results reveal that in polyoma late RNA processing splice sites can interfere with each others' usage. Cryptic splice site interference data with ß-globin IVS1 and IVS2 suggests that this may be a more general phenomenon. It is possible that splicing interference due to separation of splice sites by much less than 51 bases (9 bases in the case of ALM) generally triggers RNA degradation in the nucleus.

In vitro results have demonstrated factor interactions with 5' and 3' splice sites that correlate with splicing activity (14-17). RNAse protection experiments have revealed that specific factor - precursor RNA splice site interactions can occur whether or not all binding sites are present. For example, branch point protection has no 5' splice site requirement, while 5' splice site protection has no 3' splice site requirement in the model RNA's examined (16,17). Furthermore, the active splicing complex, or "spliceosome", has been partially purified from a HeLa splicing extract and shown to be of substantial size (50S) (18,45). A model of mRNA splicing where factors bind to the precursor RNA at specific sites and then come together into a splicing complex seems likely. This model suggests that juxtaposed splice sites with similar affinities for splicing factors can interfere with each others' ability to work. This could be at one of two levels. Two proximal sites could sterically inhibit proper factor association at either site, or factors might bind correctly but the proximity of the factor-associated splice sites sterically inhibits the coming together of 5' and 3' splice site and the branch point necessary for splicing. Such a scheme could explain the splicing defect in ALM, as depicted in our simplified model (Figure 8). The important idea is that steric hindrance blocks complete spliceosome formation, and that this leads to precursor RNA instability *in vivo*. In SLM/MP8 this defect is relieved by separating the two splice sites by an additional 42 bases. We have recently constructed a series of polyoma mutants with late leaders of various lengths, and we are in the process of analyzing these for both splicing and stability

defects. Another way to relieve this splice site interference would be to eliminate the 3' splice site (or factor binding site) at the 5'-end of the ALM late leader exon. One would predict from the model that this would restore usage of the 5' leader exon splice site. We have constructed a number of such mutants and we are in the process of analyzing them.

Our model suggests that in ALM the juxtaposition of efficient splice sites leads by steric hindrance to an inability to form functional spliceosomes. This triggers RNA degradation in the nucleus. This model does not apply to short exons where the flanking splice sites have significantly different affinities for splicing factors, or different efficiencies of splice site usage. Thus, for example, a strong 5' splice site a few nucleotides from a weak alternative (or cryptic) 3' splice site should not induce RNA degradation. This model is consistent with the recent report (46) of extremely short exons (one is only 6 nucleotides) in cardiac muscle troponin T pre-mRNA, which can be spliced in a number of alternative ways. Of course, it is also possible that the mechanism of pre-mRNA splicing for polyoma late messages is different from that of some other systems.

We have not yet examined the splicing of ALM precursor RNA in an *in vitro* splicing system. Results with other precursor RNA's have shown that crowding of splice sites *in vitro* seems to show little effect on the production of stable spliced products. The same splice sites are used *in vitro* as *in vivo* but all tend to be used at high levels (47,48). The RNA's tested are ones that do show splicing and/or stability defects *in vivo*. This leads us to speculate that the *in vitro* system may not contain the degradative system responsible for turnover of precursor RNA improperly associated with a splicing complex and so is more tolerant to splice site crowding.

Our results with S1 mapping of early RNA's using long and short probes demonstrate that a 3072 base probe gives identical results to 200-300 base probes when used to measure the relative levels of wild type and ALM RNA's. Results for late RNA, however, are different. Long and short probes detect similar levels of wild type late transcripts. With one exception (the 3' probe), however, short probes indicate 10 times more ALM RNA than long ones do. The simplest explanation of why a 337 base probe that binds across the origin (presumably indicating levels of giant RNA) shows a 40-fold ALM deficit (Figure 3), while a 3072-base probe shows a 400-fold deficit (Figure 5), is that some of the steady state RNA in ALM-infected cells is partially degraded. The ALM splicing defect quite likely leads to a different distribution of late nuclear transcripts than wild type, and might also induce an altered rate of RNA turnover. However, because the mechanism of RNA metabolism in wild type-infected cells is not well understood, a more precise explanation of the exact mechanisms of RNA processing in ALM infected cells is impossible at this time. As the short 3' probe that spans the polyadenylation site shows an approximate 100-fold deficit for ALM RNA, while a mVP1 probe of about the same size shows a 40-fold deficit, we suggest that the 3' polyadenylation region is especially sensitive to a cellular ribonuclease. As this site is a 3'-end, an exonuclease might be involved, as has been implicated

in histone mRNA turnover (49). ALM might allow us to begin to search for eukaryotic enzymes involved in RNA turnover in the nucleus.

ACKNOWLEDGEMENTS

We would like to acknowledge the excellent technical assistance of Christopher W. Marlor and Janice Seagren for help with the manuscript. This work was supported by a grant from the American Cancer Society and a State of Connecticut High Technology Research Award. G.C. is an Established Investigator of the American Heart Association.

§Present address: The Rockefeller University, 1230 York Avenue, New York, NY 10021

REFERENCES

1. Felber, B., Orkin, S. & Hamer, D. (1982) Cell 29, 895-902.
2. Montell C., Fisher,E. , Caruthers, M. & Berk, A. (1982) Nature 295, 380-384.
3. Solnick, D. (1981) Nature 291, 508-510.
4. Treisman, R., Proudfoot, N., Shander, M. & Maniatis, T. (1982) Cell 29, 903-911.
5. Treisman, R., Orkin, S.H. & Maniatis, T. (1983) Nature 302, 591-596.
6. Wieringa, B., Meyer, F., Reiser, J. & Weissmann, C. (1983) Nature 301, 38-43.
7. Wieringa, B., Hofer, E. & Weissmann, C. (1984) Cell 37, 915-925.
8. Rautmann, G. & Breathnach, R. (1985) Nature 315, 430-432.
9. Breathnach, R. & Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
10. Chu, G. & Sharp, P. (1981) Nature 289, 378-382.
11. Khoury, G., Gruss,P., Dhar, R. & Lai, C-J. (1979) Cell 18, 85-92.
12. Kuhne, T., Wieringa, B., Reiser, J. & Weissmann, C. (1983) EMBO J. 2, 727-733.
13. Solnick, D. (1985) Cell 43, 667-676.
14. Mount, S.M., Petterson, I., Hinterberger, M., Karmas, A. & Steitz, J.A. (1983) Cell 33, 509-518.
15. Black, D.L., Chabot, B. & Steitz, J.A. (1985) Cell 42, 737-750.
16. Chabot, B., Black, D.L., LeMaster, D.M. & Steitz, J.A. (1986) Science 230, 1344-1349.
17. Ruskin, B. & Green, M. (1985) Cell 43, 131-142.
18. Frendewey, D. & Keller, W. (1985) Cell 42, 355-367.
19. Acheson, N.H. (1978) Proc. Natl. Acad. Sci. USA 75, 4745-4748.
20. Treisman, R. (1980) Nucl. Acids Res. 8, 4867-4888.
21. Adami, G.R. & Carmichael, G. G. (1986) J. Virol. 58, 417-425.
22. Greenberg, M.E. & Ziff, E.B. (1984) Nature 311, 433-438.
23. Groudine, M., Peretz, M. & Weintraub, H. (1981) Molec. Cell. Biol. 1, 281-288.
24. Nevins, J.R. (1980) Methods in Enzymol. 65, 768-785.
25. Weaver, R.F. & Weissmann, C. (1979) Nucl. Acids Res. 7, 1175-1193.
26. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor,NY).
27. Maxam, A. & Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
28. Messing, J. (1983) Methods in Enzymol. 101, 20-78.
29. Feinberg, P.F. & Vogelstein, B. (1983) Analyt. Biochem. 132, 6-13.
30. Cowie, A. , Tyndall, C. & Kamen, R. (1981) Nucl. Acids Res. 10, 6305-6322.
31. Heiser, W.C. & Eckhart, W. (1982) J. Virol. 44, 175-188.
32. Tooze, J., ed. (1980) Molecular Biology of DNA Tumour Viruses (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
33. Soeda, E., Arrand, J., Smolar, N., Walsh, J. & Griffin, B. (1980) Nature 283, 445-453.
34. Treisman, R. & Kamen, R. (1981) J. Mol. Biol. 148, 273-301.
35. Ruley, E. & Fried, M. (1983) J. Virol. 47, 233-237.

36. Villareal, L.P. & White, R.T. (1983) Mol. Cell. Biol. 3, 1381-1388.
37. Acheson, N. (1984) Mol. Cell. Biol. 4, 722-729.
38. Hamer, D.H. & Leder, P. (1979) Cell 18, 1299-1304.
39. Lai, C. & Khoury, G. (1979) Proc. Natl. Acad. Sci. USA 76, 71-75.
40. Carlock, L. & Jones, N.C. (1981) Nature 294, 572-574.
41. Gallwitz, D. (1982) Proc. Natl. Acad. Sci. USA 79, 3493-3497.
42. Benyajati, C., Place, A.R., Wang, N., Pentz, E. & Sofer, W. (1982) Nucl. Acids Res. 10, 7261-7272.
43. Ruskin, B., Greene, J. & Green, M. (1985) Cell 40, 833-844.
44. Busslinger, M., Moschonas, N. & Flavell, R. (1981) Cell 27, 289-298.
45. Grabowski, P.J., Seiler, S.R. & Sharp, P.A. (1985) Cell 42, 345-353.
46. Cooper, T.A. & Ordahl, C.P. (1985) J. Biol. Chem. 260, 11140-11148.
47. Krainer, A.R., Maniatis, T., Ruskin, B. & Green, M. (1984) Cell 36, 993-1005.
48. Reed, R. & Maniatis, T. (1985) Cell 41, 95-105.
49. Ross, J., Peltz, S.W., Kobs, G. & Brewer, G. (1986) Mol. Cell. Biol. 6, 4362-4371.